



Earth System Grid

The Earth System Grid Center for Enabling Technologies (ESG-CET): Community Data Resource Evolution

**R Ananthakrishnan, D E Bernholdt, S Bharathi, D Brown, M Chen, A L Chervenak,
L Cinquini, R Drach, I T Foster (PI), P Fox, D Fraser, K Halliday, S Hankin, P Jones,
C Kesselman, J D E Middleton (PI), J Schwidder, R Schweitzer, R Schuler, A Shoshani,
F Siebenlist, A Sim, W G Strand, N Wilhelmi, M Su, D N Williams (PI)***

Summary

The Earth System Grid Center for Enabling Technologies (ESG-CET)—part of the Department of Energy’s SciDAC-2 investment in climate change research—addresses U.S. and international management and sharing of distributed petascale data. Its goals are to sustain the successful existing ESG system, address future scientific needs for data management and analysis, extend ESG to support sharing and analysis of climate simulation data from the Intergovernmental Panel on Climate Change (IPCC) 5th Assessment Report (AR5) in 2010, support the Climate Science Computational End Station (CCES) at the DOE Leadership Computing Facility at Oak Ridge National Laboratory (ORNL), and support climate model evaluation activities under other SciDAC-2 projects, in particular “A Scalable and Extensible Earth System Model for Climate Change Science.” To achieve our goals, the ESG-CET leads and participates in global efforts to standardize methods used to access and analyze federated data. It is our intent that the approaches and technologies developed in ESG-CET also impact other SciDAC-2 scientific application areas, such as astrophysics, molecular biology, and materials science.

Introduction

The Community Climate Simulation Model (CCSM), CCES, IPCC, and North America Regional Climate Change Assessment Program (NARCCAP), and other national and international programs have entrusted their current and future data archives and distribution to ESG.

To meet these challenging demands of these large and diverse user populations, we developed metadata technologies (standard organization, metadata extraction based on netCDF, and Metadata Catalog Services), security technologies (web-based user registration and authentication, and community authorization), data transport

technologies (for high-performance access, robust multiple file transport, integration with mass storage systems, and support for dataset aggregation and subsetting), and web portal technologies (to provide interactive access to the data holdings). We also integrated support for popular analysis and visualization tools.

ESG’s success in disseminating climate data has exceeded all expectations, and ESG is recognized for its critical role in U.S. and international climate research.

Technology Impact in 2007

We finalized the first version of the *ESG-CET domain model*, which provides a

* (925) 423-0145, williams13@llnl.gov

logical conceptualization of the objects and relationships needed to support next-generation ESG data services. This model encompasses science metadata (collection-level, inventory-level, and item-level), user management, access control, and metrics reporting.

We have also devoted considerable effort to investigating the role that emerging semantic technologies (e.g., RDF, OWL, Sesame) can play in next-generation ESG data search and discovery services. We developed prototype search services and interfaces against current IPCC, CCSM, and Parallel Climate Model (PCM) metadata, in order to test performance, flexibility and scalability of this approach. Initial results are encouraging, and work continues.

Climate Community Impact in 2007

Data delivered via the two ESG portals since December 2004 has supported the authoring of more than 300 papers. This rapid scientific return is unprecedented in the climate community and attests to the ease with which users were able to retrieve model output from ESG.

Virtually all of the new model results reported on in the 2007 IPCC's 4th Assessment Report (AR4) Working Group 1 was based on the ESG IPCC archive. The simulation data consists of 12 experiments, each performed by 24 models from 17 climate modeling centers in 13 nations, located on 4 continents.

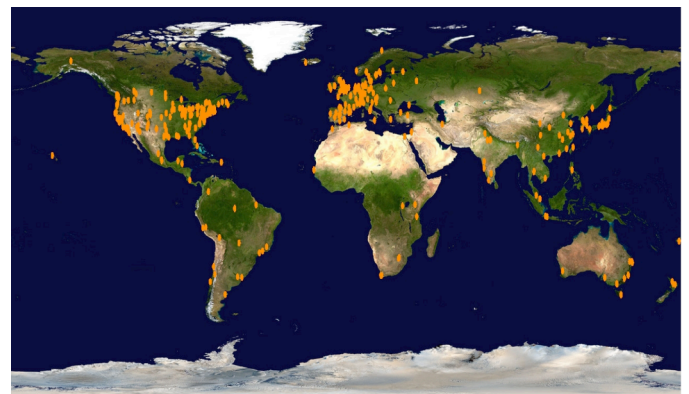
Equally important is the ESG CCSM data archive, which supplies additional climate model output, including PCM, Parallel Ocean Program, Land, and Sea Ice.

In aggregate, ESG-CET provides access to distributed climate simulation data to over 6,000 registered users worldwide (figure), who have downloaded more than 250 terabytes. The archives contain more than

100,000 simulated years of data: over 1 million files and more than 180 terabytes.

We have established within ESG support for publishing and distributing NARCCAP data, and developed an extensive data management plan that involves distributed data access from the ESG CCSM portal at NCAR to data resources at NCAR and LLNL. We have approved the first test users for access.

We also established a portal at ORNL in support of the Climate Science Computational End Station project. This portal will allow delivery of data to users directly from its primary storage site at ORNL, and anticipates the second-generation federated architecture currently under development by the ESG team, which will support delivery of data to users with minimal extra data movement. This portal will initially support the CCSM Climate Land Model Intercomparison Project (C-LAMP), which is integrating biogeochemical models into CCSM.



ESG Usage Internationally

For further information contact:

Dean N. Williams, LLNL

williams13@llnl.gov, (925)-423-0145; or

Don Middleton, NCAR

don@ucar.edu, (303) 497-1250; or

Ian Foster, ANL

foster@mcs.anl.gov, (630) 252-4619